A/B TESTING

RADITYA ERLANG ARKANANTA

March 2025

ABOUT DATASET

This ad-ab-testing.csv dataset is a dataset from Kaggle (https://www.kaggle.com/datasets/osuolaleemmanuel/ad-ab-testing), where in this dataset, there are a number of users divided into 2 groups, namely the

- Control Group that will see the Ad Basic Template, and the
- Exposed Group that will see the SmartAd Brand Interactive Ad.

Objectiev

Conduct an analysis of both groups to find out whether SmartAd is successful, and whether there is a significant difference between the two groups, and find out the sample size and A/B testing period that should be carried out.

HYPOTHESIS

- **Null Hypothesis (H_0):**

There is No Significant Impact of SmartAd on Conversion

- **Alternative Hypothesis (H_1)**:

There is Significant Impact of SmartAd on Conversion

DATA UNDERSTANDING

- **auction_id**: Unique ID of the User who has been presented with the BIO Questionnaire.

  If you view it but do not answer Yes or No, then both Yes and No will be Zero.
- **experiment**: Group User, Exposed or Control
- **control**: Users Viewing Ad Standard Template
- **exposed**: Users Viewing SmartAd Brand Interactive Ads.
- **date**: Date (YYYY/MM/DD)
- **hour**: Hour.
- **device_make**: Device used for access
- **platform_os**: OS ID.
- **browser**: Name of Browser Used.
- **yes**: 1 If User Answers Yes on BIO Questionnaire.
- **no**: 0 If User Answers No on BIO Questionnaire.

DATA PRE-PROCESSING

```
# Cek Data
df.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8077 entries, 0 to 8076
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   auction_id   8077 non-null   object
 1   experiment   8077 non-null   object
 2   date         8077 non-null   object
 3   hour         8077 non-null   int64
 4   device_make  8077 non-null   object
 5   platform_os  8077 non-null   int64
 6   browser      8077 non-null   object
 7   yes          8077 non-null   int64
 8   no           8077 non-null   int64
dtypes: int64(4), object(5)
memory usage: 568.0+ KB
```

```
### Mengecek Tidak Ada
df.isnull().sum()
```

[65] ✓ 0.0s

```
auction_id     0
experiment     0
date           0
hour           0
device_make    0
platform_os    0
browser        0
yes            0
no             0
dtype: int64
```

- **Check Missing Value**

There is no Missing Value

```
### Drop Duplicate
len(df.drop_duplicates())/len(df) # 1, TIdak Ada Duplikat
```

✓ 0.0s

```
1.0
```

- **Check Duplicate Value**
- There is no Duplicate

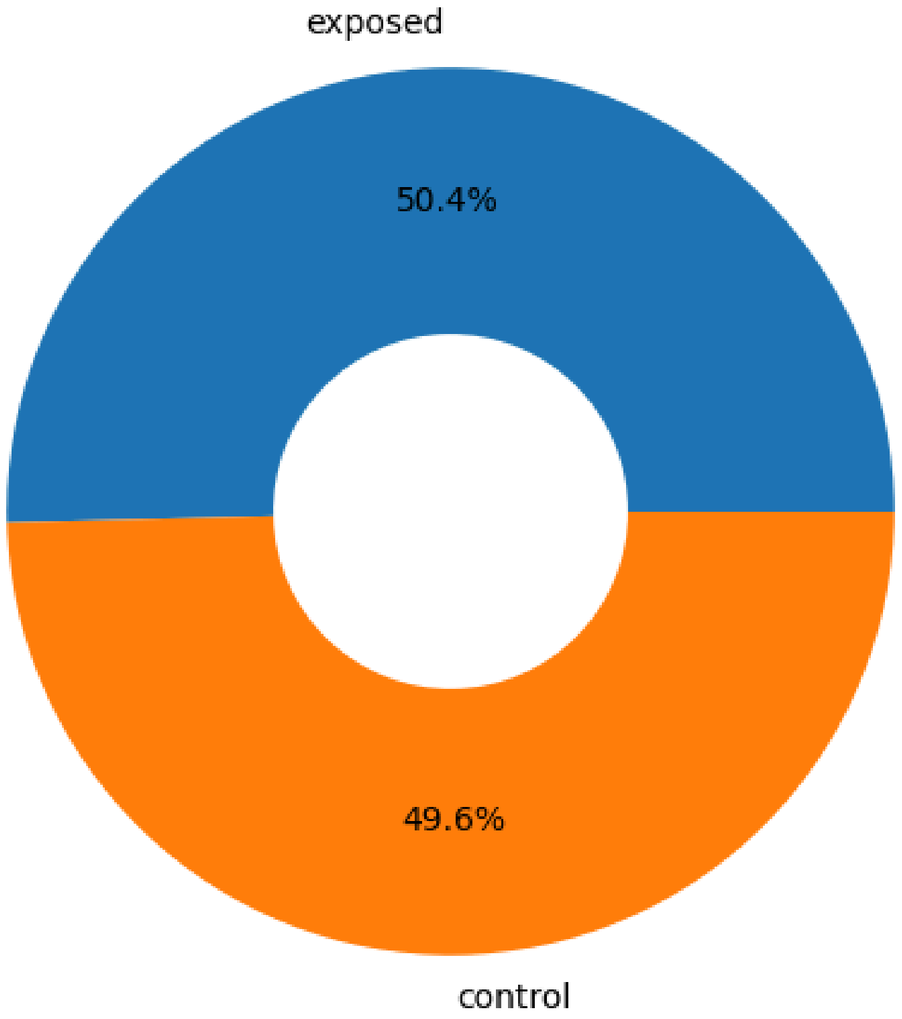
DATA PROCESSING

```
# Menghitung Jumlah Tiap Group dan Persentasenya
print(df['experiment'].value_counts())
df['experiment'].value_counts(normalize=True)
```

[67] ✓ 0.0s

```
experiment
control    4071
exposed    4006
Name: count, dtype: int64

experiment
control    0.504024
exposed    0.495976
Name: proportion, dtype: float64
```



Number of Each Group and Percentage

From 8077 Data, It Was Found That

- Control Group 4071 at 50.4%
- Exposed Group 4006 at 49.6%

DATA PROCESSING

```
# Cek Rata - Rata Yang menekan Ad di Kedua Group
df.groupby('experiment')['yes'].mean()
```

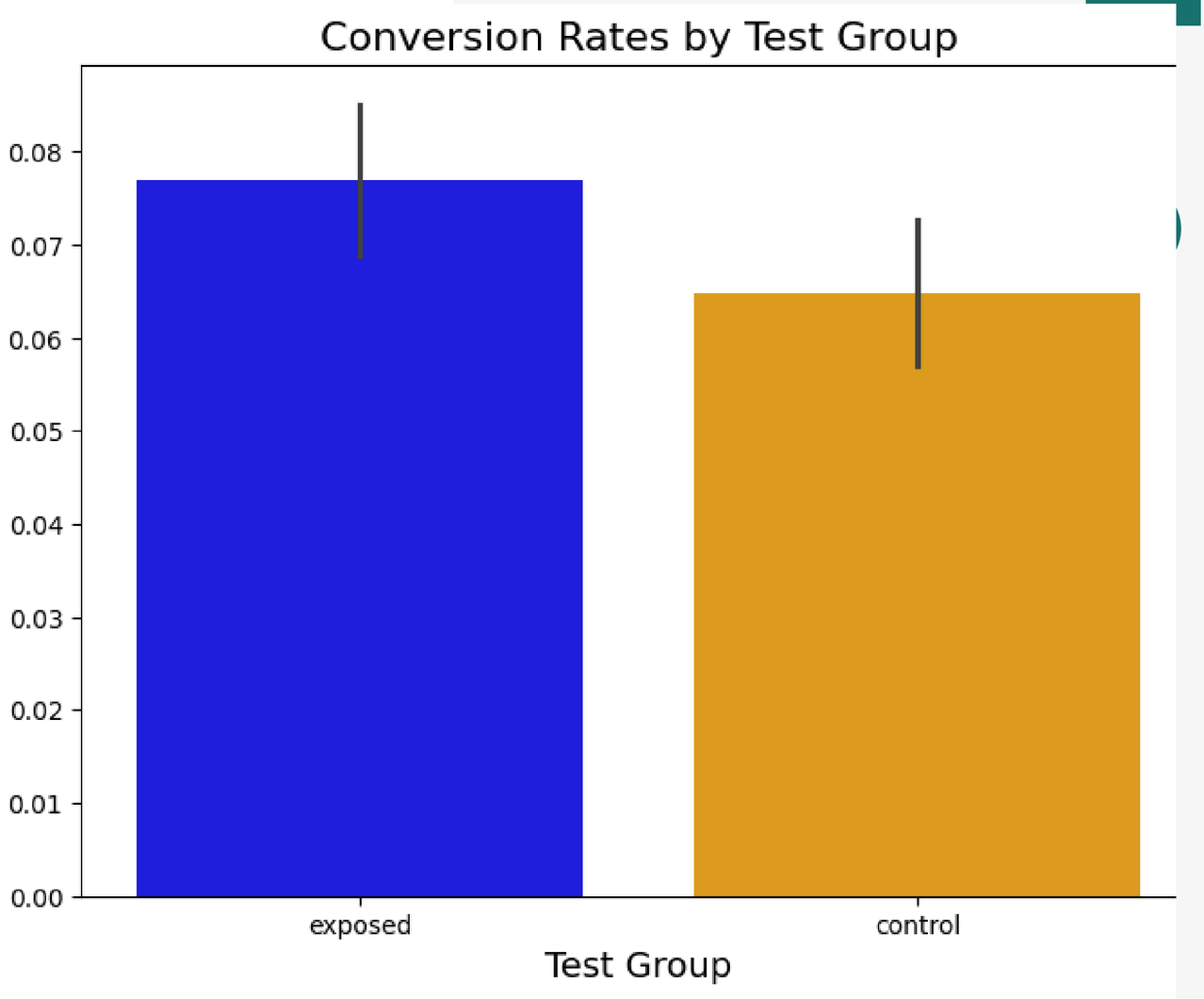
✓ 0.0s

```
experiment
control    0.064849
exposed    0.076885
Name: yes, dtype: float64
```

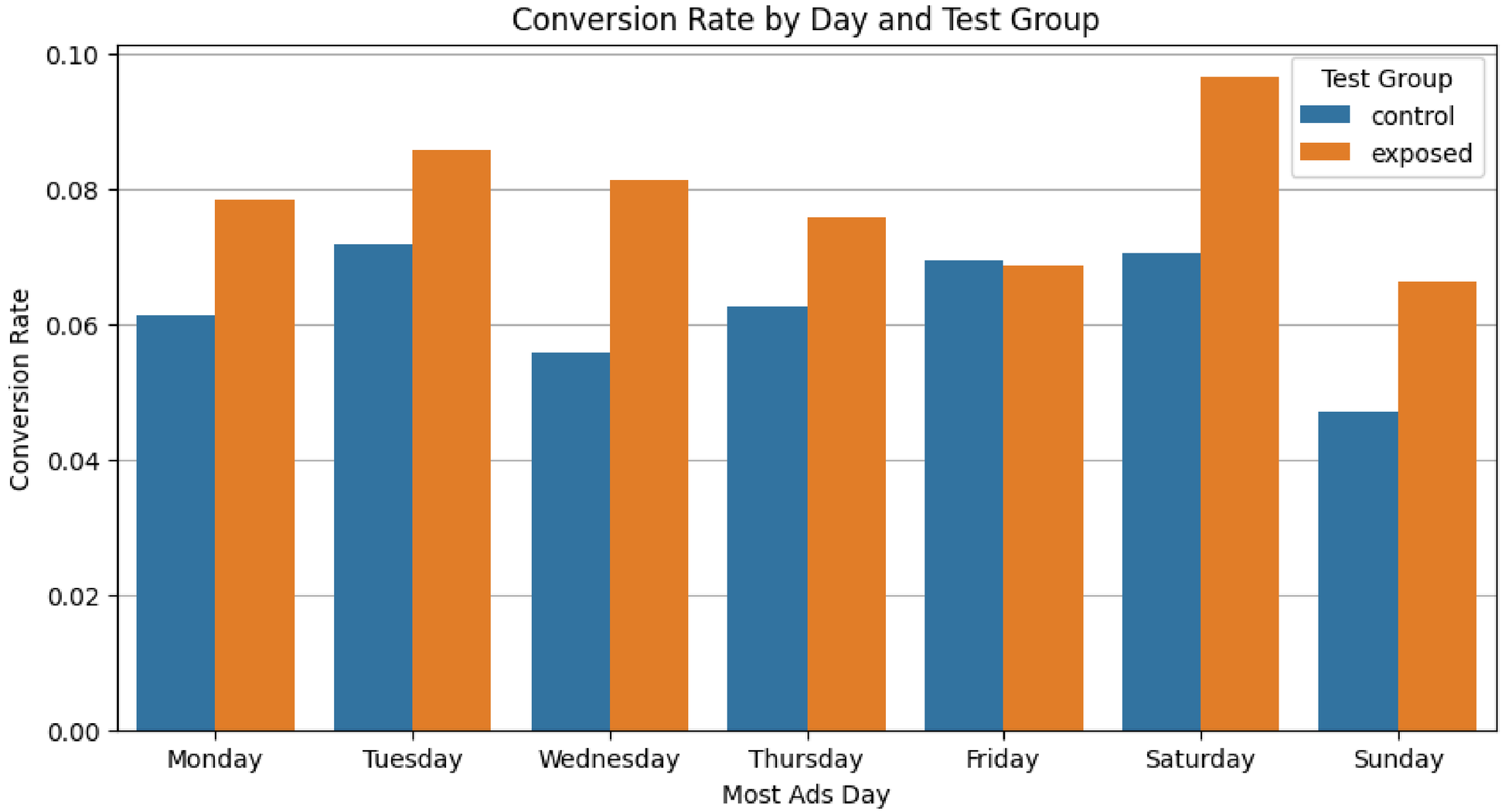
Average Users Who Clicked on Ad for Each Group (Conversion Rate)

It was found that the Conversion Rate

- Control Group is 6.48%
- Exposed Group is 7.68%



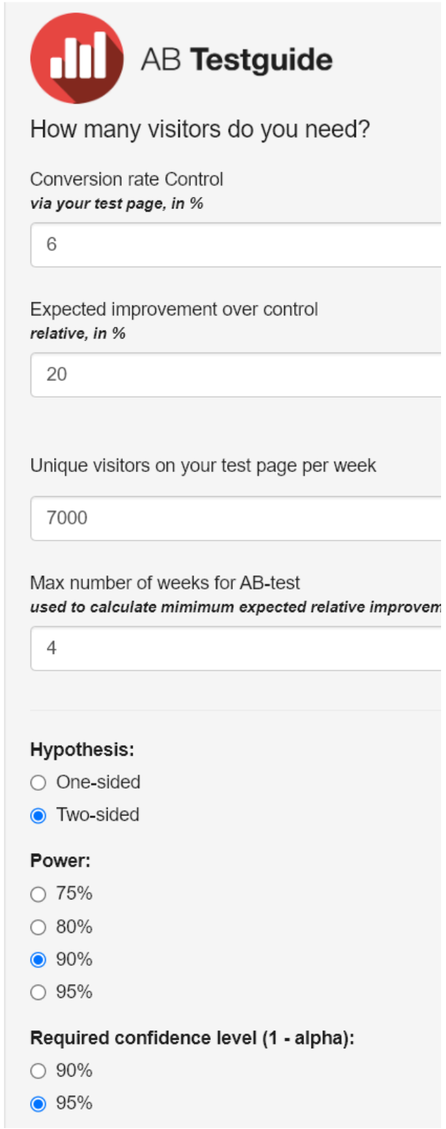
CONVERSION RATE BY DAY AND GROUP

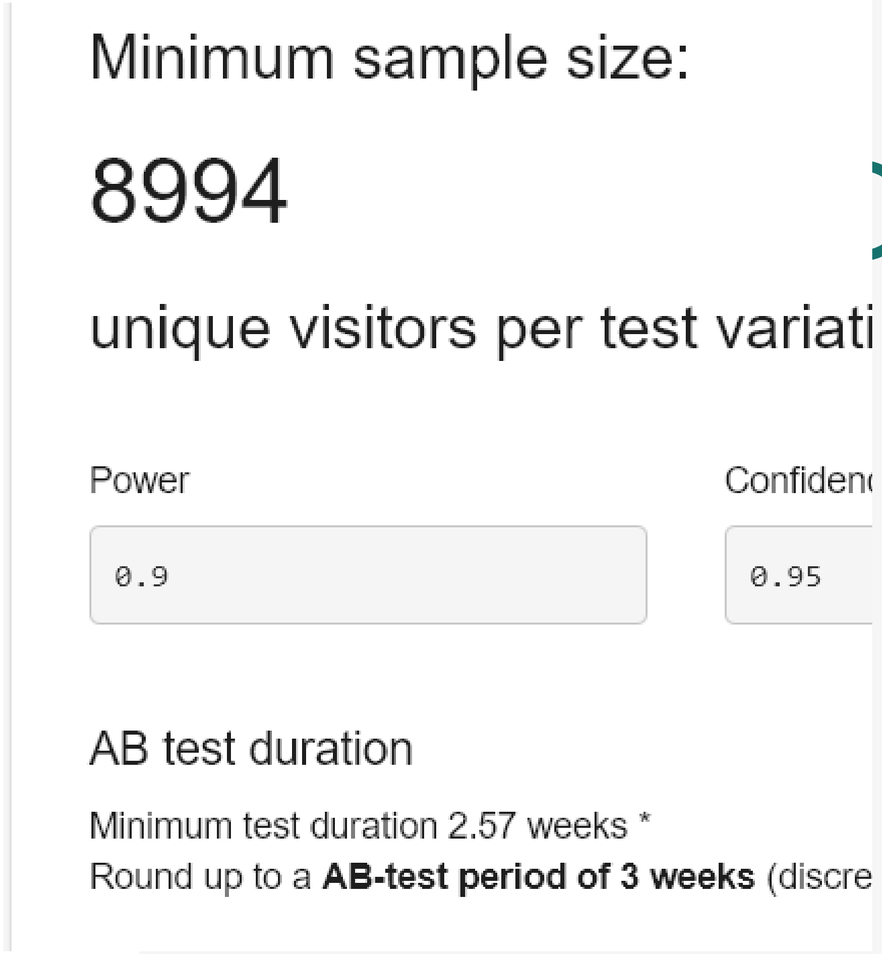


SAMPLE SIZE & DURATION

Sample Size and Duration

- In the recorded data, User arrivals for 8 Days, where in those 8 Days there were 8077 Visitors, so it is assumed that there will be 1000 Visitors each day
- Using ABtestguide with Power 10%, Control Conversion Rate 6% with Expected Improvement 20% from 6%, Unique Visitors per week 7000
- A Sample Size of 8994 was obtained with a Duration of 2.57 Weeks, so to validate the results of this A/B Testing, a Sample Size of 8994 with a Duration of 3 Weeks (Rounded Up) is required





T-TEST

```
# Nilai untuk T-Test, yang menjawab Yes di Kedua group
control_group = df[df['experiment'] == 'control']['yes']
exposed_group = df[df['experiment'] == 'exposed']['yes']
```

[71] ✓ 0.0s

```
# T-Test
t_stat, p_value = ttest_ind(control_group, exposed_group, equal_var=False)
```

[72] ✓ 0.0s

```
# Menunjukkan Hasil T-Test
print(t_stat)
print(p_value)
```

[73] ✓ 0.0s

```
-2.107278649715408
0.035124460220699455
```

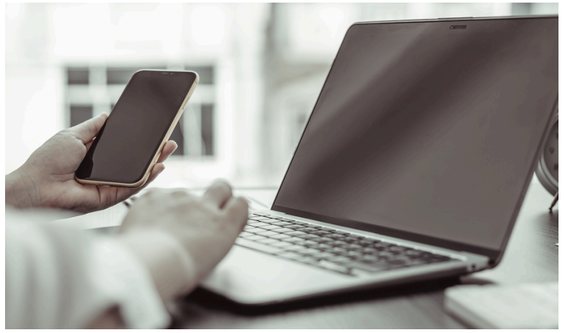
- **Because P < 0.05**

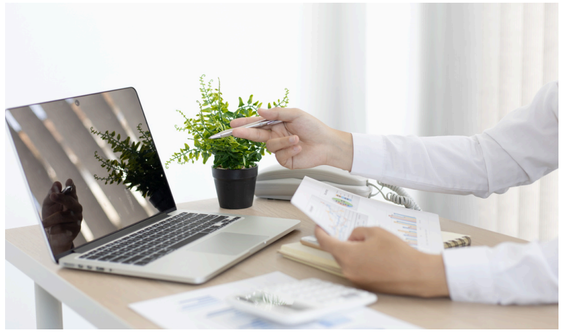
There is a Significant Difference Between the Group Exposed to SmartAd and the Control Group

Meaning that Users Who Get SmartAd Will Tend to be More Interested in Answering Ads and Converted

INSIGHT

SOLUTIONS OF THE PROBLEMS

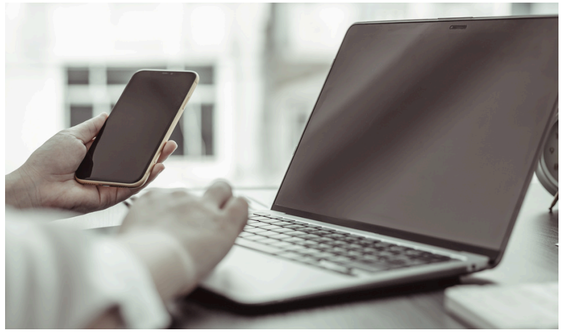


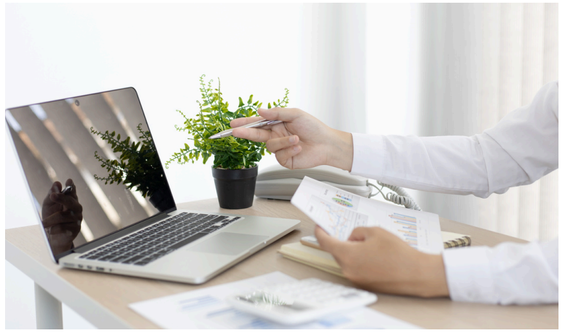


- It can be seen that the group that received SmartAd had a significantly higher conversion rate compared to the group that only received Basic Ad except on Friday where the Control Group gave a slightly higher Conversion, but on other days the Exposed Group consistently gave a higher Conversion Rate.
- In addition, it was found that the most Conversions were obtained on Saturdays for both the Control Group and the Exposed Group.
- Thus, companies can allocate more advertising budgets on those days to maximize conversions.
- Companies can also optimize their marketing strategies by adjusting the ad schedule during the hours with the highest traffic and testing various ad variations to increase campaign effectiveness.

RECOMMENDATION

SOLUTIONS OF THE PROBLEMS





- Increase the frequency of Ad views on Saturdays when engagement is highest.
- Placing Ads in strategic locations within the app that users visit more often on those days.
- Implementing dynamic Ads that change based on the time of day, preferences, or user activity.
- Sending push notifications that direct users to Ads at the right time.
- Analyzing interaction patterns with heatmaps or session tracking to see if Ad placement is optimal.

Thynk Unlimited

THANK YOU

FOR YOUR NICE ATTENTION